LeaRning R in ChemistRy at Reed College

Authors: Danielle Cass[1] and Chester Ismay[2]

1. Reed College, Portland, OR. 2. Data Camp, https://www.datacamp.com/

Abstract

In 2015 general chemistry at Reed made the decision to switch from a spreadsheet system to using a more comprehensive program R and the RStudio interface to perform all data analysis and visualization. Excel was originally used for this purpose and the advantage Excel had was that most students had experience using it in high school. Therefore there was not a steep learning curve for its use in labs. However, the switch was motivated by a continued frustration with the limits of Excel and a newly acquired R expert on the Reed staff. In hindsight we can also add that this switch is advantageous to the student's future. This paper provides an outline of why and how we implemented the use of R in the introductory curriculum.

Introduction

Educational ideas are constantly a swinging pendulum. Teaching computer skills is no different. In 1988 a plenary speaker at the BCCE conference explained how departments should shift their computer focus from coding to software skills (1). Yet in 2018 the opposite was suggested, as an editorial in C&EN news said that the future of chemistry will include directly working and talking with machines (2). The author goes on to say, "Unfortunately, few chemists can actually code, let alone program a robot or write an algorithm" (2). If we were trying to provide students an education that they could take with them, a switch to R seems like a solid move; feeding a need in our current and future chemistry work force.

So why the switch from Excel to R, how are they different and what does R offer that Excel does not? R is a programming language that was originally developed by two statisticians in New Zealand (3). The way data analysis is executed and visualized in R is much different than Excel. First of all, everything done in R is written in code instead of the traditional point and click format of spreadsheets. RStudio, the interface that is being used, shows the code and the results of the analysis separately (Figure 1A). It is also possible to directly integrate the code into your final report (Figure 2). This is drastically different from Excel which does not use a programming language but instead displays either the formula used or the data and there is no record of the changes made by the point and click options (Figure 1B). This leads to the second advantage of R, reproducibility. To reproduce an analysis in Excel, it is often complicated and requires many changing variables. R, however, allows a person to set up code for creating a scatter plot fit to a straight line and then anyone who wants to create this type of plot can use the same code. This is an

important point for general chemistry because it allows students to begin creating an analysis tool box which they can take with them anywhere.

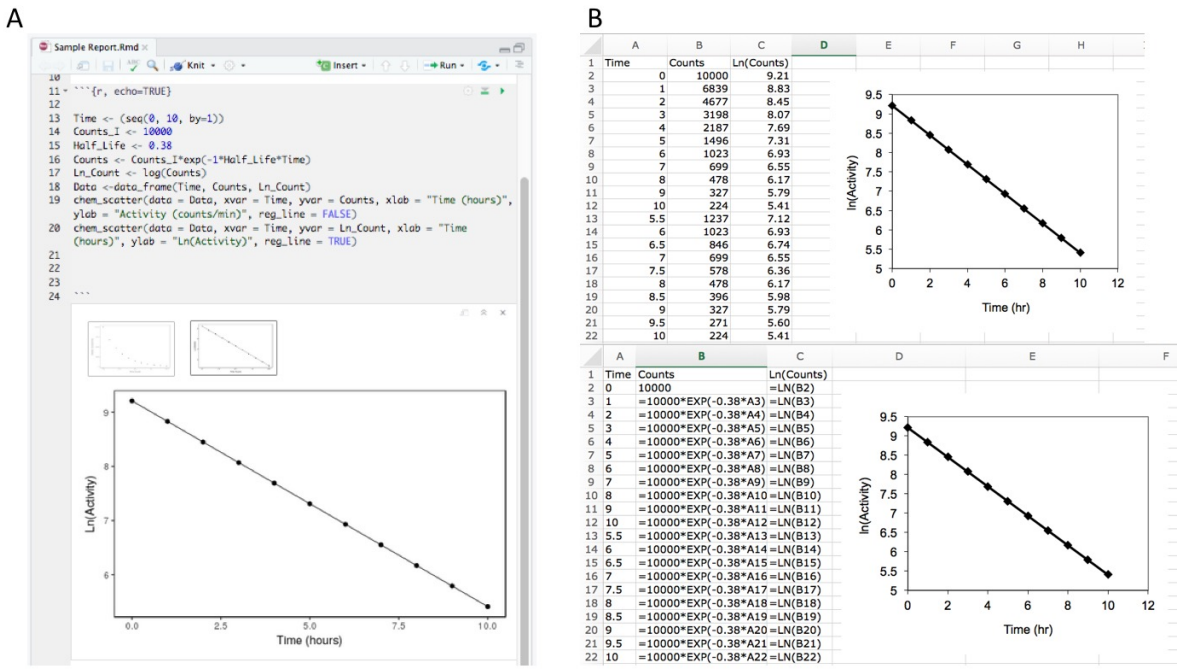


Figure 1: The RStudio interface compared to the Excel interface. A. Every step made to create the graph in R is written in the code lines 13-20 with the resulting graph shown below. B. Excel has no record of the graphical changes and mathematical changes can only be visualized separately (bottom).

Sample Report

```
Time <- (seq(0, 10, by=1))
Counts_I <- 10000
Half_Life <- 0.38
Counts <- Counts_I*exp(-1*Half_Life*Time)
Ln_Count <- log(Counts)
Data <-data_frame(Time, Counts, Ln_Count)
chem_scatter(data = Data, xvar = Time, yvar = Counts, xlab = "Time (hours)",
ylab = "Activity (counts/min)", reg_line = FALSE)

chem_scatter(data = Data, xvar = Time, yvar = Ln_Count, xlab = "Time
(hours)", ylab = "Ln(Activity)", reg_line = TRUE)
```

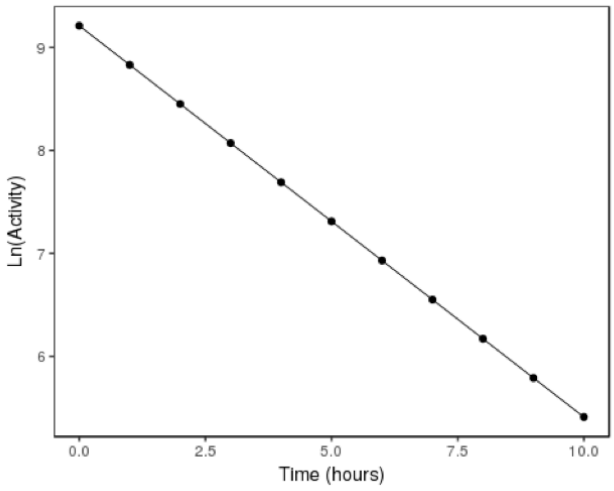


Figure 2: Output of RStudio into a Word document. The code used to create the graphs can be shown in the final document as boxed areas.

Excel is also limited in its data set size while R can quickly handle an infinitely large dataset. Although this is not relevant in most general chemistry analysis, it is important when thinking about their future and what kind of analysis they are likely to do. Finally, one important factor to remember is that R is open source and free. Excel requires a license to use, which means that there is a monetary hurdle in its use. Excel is provided, free, to students at Reed, but that is limited to their time as students. Also, because of the open-source format there is a large user group on-line that has created packages that may be useful to specific types of analysis.

Introducing R

As mentioned above, a hurdle to using R for data analysis is that most students don't start the course with any experience using the language and program. Therefore we needed to change the lab structure to allow for more in-class analysis time. This is good for providing help to students regarding R usage, but it also just a good pedagogical plan. Most students don't appear to struggle with the acquisition of data, but what to do with that data is where they need extra help. Yet, traditionally the labs had an instructor and TA present during all of the acquisition stage and very little presence during the analysis

stage. The change to R made us rethink this strategy. The result is that we went from sixteen to twelve labs during the entire first year lab sequence, spreading labs into multiple weeks and using the extra time for data analysis. This allowed us to spend more time helping students with R and their subsequent analysis during their allotted lab sections.

Currently, the second lab of the semester is where R is initially introduced. The data acquisition stage occurs during week three of the semester and then the data analysis occurs the following lab period. Previously we had tried to begin with R in week 1, but found that introductory students are mostly freshmen and it was too much for students to comprehend along with everything else happening during their first weeks as a college student. In this first analysis period we introduce students to the RStudio platform, introduce them to the resources they will have available to them during all subsequent lab reports, and help them through their data analysis. This is done through a short lecture and then a step-by-step instruction guide (see attached pdf).

How to introduce students to R without devoting an entire course to it is a continuing struggle for us. The rest of the paper will be devoted to explaining the resources we have developed and used to help in this regard.

Chemistr

One advantage to R, as mentioned above, is the open-source aspect. This allows many people to develop their own code and share it with others. In R these shareable sets of code are called packages and can be accessed from a variety of locations. The main location is CRAN (Comprehensive R Archive Network). Currently there are over 12,000 packages available which means that many computational scenarios have already been worked out and have gone through rigorous quality checks (4). However, we decided to create our own package. The advantage to this is that we could make it specific to the needs and abilities of the introductory chemistry students. This package is called Chemistr and is available through github, another location for publishing packages which have not gone through the quality checks required by CRAN (5). The Chemistr package is designed to assist students with little to no programming experience create reproducible analysis using R and R Markdown. Currently there are six defined functions in Chemistr (Figure 3a), one of which is chem_scatter. This function takes a complex set of code (Figure 3b) and simplifies it into a small set of arguments that students can see using the HELP window in RStudio(Figure 3c). These arguments are relative to the different types of plots the students will be creating (Figure 3d).

A
Facilitates use of R functions and R Markdown files for Chemistry lab reports
Documentation for package 'chemistr' version 0.0.3
DESCRIPTION file.
Help Pages
chem_clustered.bar Creates a clustered bar graph
chem_dual.scatter Creates a scatterplot with a common horizontal axis and two different variables plotted on the vertical axis with regression lines
chem_lab_word Creates an R Markdown Word document for the Reed College Chem 101/102 labs
chem_scatter Creates a scatterplot with regression line
chem_table Creates a nice table for data frames
periodic_table Data on the elements of the periodic table
B
C
Files Plots Packages Help Viewer
chem_scatter {chemistr} R Documentation
Creates a scatterplot with regression line
Description
Creates a scatterplot with regression line
Usage
Arguments
D
Scatter Plot

Figure 3: Chemistr package. A. The functions that are part of the chemistr package. B. The code that is behind the use of the chem_scatter function. C. The information available through the HELP window about the chem_scatter function. D. The use of the chem_scatter function to create and analyze a plot of transmittance versus concentration. This is a plot that is created for the first R Studio lab.

Templates

Another advantage to creating our own package is that when students use Chemistr, they also gain access to all the templates created specifically for every lab report that uses R. These templates provide a basic layout of their report along with helpful hints and code when necessary (Figure 4a). Students then fill out the template by adding their own data (Figure 4b, lines 18-22), doing any mathematical manipulation of the data (Figure 4b, lines 23-27), and a statistical analysis (Figure 4b, lines 41-44). The output of this code is a word document that now has a consistent style and can be further modified for significant figures and other small changes (Figure 4c).

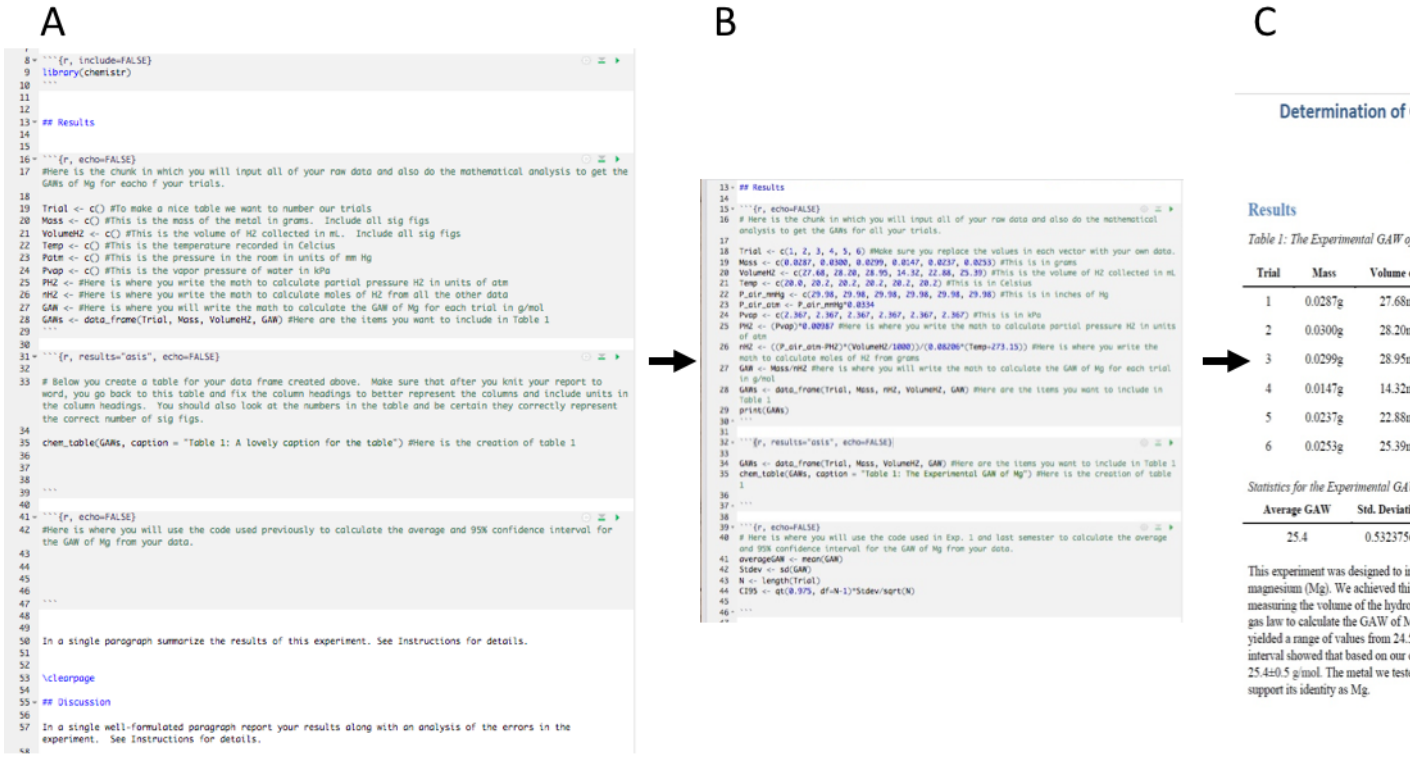
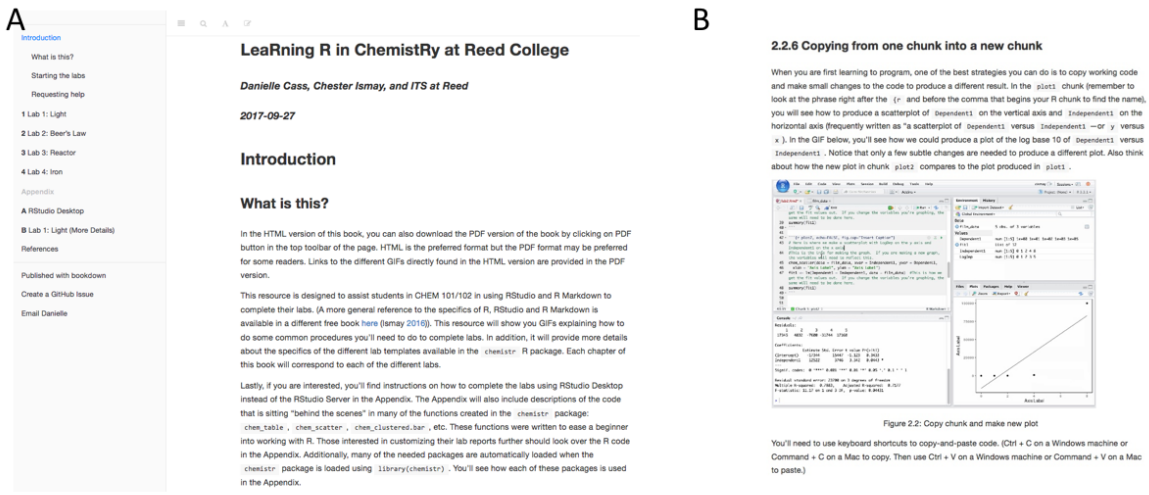


Figure 4: An example template found in the Chemistr package. A. An empty template has a outline for the report, including the types of analysis that needs to be done. B. Students then fill in the template and add extra code and formatting. C. The final report is knitted into a Word document that can then undergo final editing and modifications.

E-Book

Finally, we have created an on-line resource (6). This resource takes students through the initial setup phase of RStudio all the way to the details of writing each individual lab report. The advantage of using an e-book format is that we can include videos and gifs of how students should navigate RStudio correctly and links to relevant outside resources (Figure 5). The e-book becomes the one-stop for all lab report needs. The downside to the e-book is that it requires constant updating as labs, lab instructions, or teaching methods change. This has been difficult to do and we are in the process of making this a more sustainable endeavor.


A
Introduction
What is this?
Starting the labs
Requesting help
1 Lab 1: Light
2 Lab 2: Beer's Law
3 Lab 3: Reactor
4 Lab 4: Iron
Appendix
A RStudio Desktop
B Lab 1: Light (More Details)
References
Published with bookdown
Create a GitHub Issue
Email Danielle
LeaRning R in ChemistRy at Reed College
Danielle Cass, Chester Ismay, and ITS at Reed
2017-09-27
Introduction
What is this?
In the HTML version of this book, you can also download the PDF version of the book by clicking on PDF button in the top toolbar of the page. HTML is the preferred format but the PDF format may be preferred for some readers. Links to the different GIFs directly found in the HTML version are provided in the PDF version.
This resource is designed to assist students in CHEM 101/102 in using RStudio and R Markdown to complete their labs. (A more general reference to the specifics of R, RStudio and R Markdown is available in a different free book here (Ismay 2016)). This resource will show you GIFs explaining how to do some common procedures you'll need to do to complete labs. In addition, it will provide more details about the specifics of the different lab templates available in the chemistr R package. Each chapter of this book will correspond to each of the different labs.
Lastly, if you are interested, you'll find instructions on how to complete the labs using RStudio Desktop instead of the RStudio Server in the Appendix. The Appendix will also include descriptions of the code that is sitting "behind the scenes" in many of the functions created in the chemistr package: chem_table , chem_scatter , chem_clustered.bar , etc. These functions were written to ease a beginner into working with R. Those interested in customizing their lab reports further should look over the R code in the Appendix. Additionally, many of the needed packages are automatically loaded when the chemistr package is loaded using library(chemistr) . You'll see how each of these packages is used in the Appendix.
B
2.2.6 Copying from one chunk into a new chunk
When you are first learning to program, one of the best strategies you can do is to copy working code and make small changes to the code to produce a different result. In the plot1 chunk (remember to look at the phrase right after the {r and before the comma that begins your R chunk to find the name), you will see how to produce a scatterplot of Dependent1 on the vertical axis and Independent1 on the horizontal axis (frequently written as "a scatterplot of Dependent1 versus Independent1 —or y versus x). In the GIF below, you'll see how we could produce a plot of the log base 10 of Dependent1 versus Independent1 . Notice that only a few subtle changes are needed to produce a different plot. Also think about how the new plot in chunk plot2 compares to the plot produced in plot1 .
Figure 2.2: Copy chunk and make new plot
You'll need to use keyboard shortcuts to copy-and-paste code. (Ctrl + C on a Windows machine or Command + C on a Mac to copy. Then use Ctrl + V on a Windows machine or Command + V on a Mac to paste.)

START

Figure 5: LeaRning R in ChemistRy at Reed College. A. The book gives basic information on how to get started using R. B. There also individual chapters on each lab report. In this example there is information on how to use code from a previous lab in a new report. C. A GIF showing a user how to upload a picture that a student may want to include in their lab report.

Future Work

Now that we are starting to see how to implement RStudio into the classroom, our next project is to make it even easier for students to begin this process outside of the classroom. This entails condensing all of our lab resources, including the lab instructions, into our on-line book. Within this book we would like to include more videos and links to our other sources. Also, we are working on creating a course that mimics the style of Data Camp, specifically focusing on teaching students the basic R skills they will need for their first year of chemistry. This way students could take this course before, during, or after they have completed general chemistry.

References

1. Earl, B.L.; Emerson, D.W.; Johnson, B.J.; Titus, R.L. Teaching Practical Computer Skills to Chemistry Majors. *J Chem Ed*. **1994** *71*, 1065.
2. Martinez, J.G.; The new chemist. *C&EN News*. **2018**, 92, 2.
3. Ihaka, R.; R: Past and Future History. https://cran.r-project.org/doc/html/interface98-paper/paper.html (Accessed March 25, 2018).
4. CRAN. https://cran.r-project.org/(Accessed April 2nd, 2018).
5. Chemistr. https://github.com/ismayc/chemistr. (Accessed April 2nd, 2018)
6. Cass, D.; Ismay, C. LeaRning R in ChemistRy. https://ismayc.github.io/chemistr-book/.